myReader: an Auto-Reader for Low Vision Users

Phillip Pearson, Paul Seakins, David Lee, Nigel Tucker, and Darryl Sherwood

HumanWare Ltd
Christchurch, New Zealand
pp-das@myelin.co.nz, paul.seakins@humanware.com, david.lee@myrealbox.com,
{nigel.tucker, darryl.sherwood}@humanware.com

Abstract. Until recently, video magnifiers – closed-circuit television systems allowing magnification and colour modification – were the most effective reading aid for the visually impaired. Humanware Ltd. has developed a practical portable consumer device, myReader, incorporating a high resolution digital image sensor and document analysis software capable of understanding complex Manhattan layouts. In addition to typical video magnifier functionality, myReader can capture and analyse a full page of text and present it to the user in a more convenient format: in a single tall column, in a single long line, or one word at a time. Users in a study comparing myReader with a conventional video magnifier achieved a moderate increase in reading speed and reported a substantial increase in physical comfort when using myReader.

1 Introduction

Low vision is defined as a condition where ordinary eyeglasses, lens implants or contact lenses cannot provide sharp sight. Low vision can be caused by a variety of eye problems, each of which results in different visual characteristics; some common examples are macular degeneration, diabetic retinopathy, inoperable cataracts, and glaucoma. Individuals with low vision find it difficult, if not impossible, to read small writing or to discern small objects without high levels of magnification. This limits their ability to lead an independent life.

The most common electronic viewing device among low vision users is the closed-circuit television (CCTV) based "video magnifier". This device uses a video camera, zoom lens and monitor to view a book, document or other small object. The user can adjust the magnification of the zoom lens so they can use their remaining sight to its best advantage when viewing very small objects or writing. The picture on the monitor is displayed in real time so the user can read by physically moving the reading material underneath the camera.

The first video magnifier, developed by the Rand Corporation, was demonstrated in 1968, and the concept has not changed substantially in the intervening years. While a video magnifier provides a perfectly adequate method for reading a document, there are some disadvantages with the system. The process of scanning a book or document backwards and forwards underneath the camera takes a lot of mental concentration and physical skill, and is distracting for the user, who simply wants to read the book. The visual effect of the continuous scanning movement can also cause motion sickness. People thus do not tend to use video magnifiers to read for long periods.

In 2001, Humanware Ltd. (previously Pulse Data International Ltd.) began to develop a reading aid product without these disadvantages. The first model, myReader 600, shown in Fig. 1, was released in November 2004. As well as offering a live video function like a video magnifier, it is also able to rapidly capture an A4-sized (210 x 300 mm) image of a document using an embedded digital image sensor. The page image is analysed to extract and group text elements (characters) and other components such as pictures and formatting. Text elements are then associated into words, lines and blocks. Once the page is analysed, the text elements can be magnified and rearranged into suitable formats for the user to read, such as in a column, a line or one word at a time, and a keypress initiates automatic scrolling through the text. In this way the user can concentrate purely on reading, free from the distracting need to physically move the book backwards and forwards under the camera.

Development of myReader was a significant undertaking for Humanware, involving many engineers over several years. The myReader system architecture involves a number of components that were not readily available at the time. The document camera is based on a custom-designed multi-megapixel image sensor. The hardware uses a two-processor architecture: a general purpose CPU was chosen for document segmentation, image rendering and some image processing, while a separate processor handles the user interface and system control. A mixture of commercial video processing ICs and programmable logic provides image capture, processing and the live video mode. The whole unit is designed to fold into a portable package, allowing users to easily carry it wherever it is needed.

Fig. 1. Humanware myReader 600.

After investigating the cost and performance of existing OCR and document analysis software, the decision was made to write a complete end-to-end document analysis engine. Since it is extremely difficult to create a document analysis engine that will correctly handle arbitrary page layouts, it was decided to optimize the system to work with complex Manhattan-style layouts.

Prior to the development of myReader, there were other low vision viewing systems that also allowed the user to capture an image and to segment and reformat the text (e.g., the Horizon, from Mentor O&O, Norwell, MA). However, these systems used flat-bed scanners, which take 15-60 seconds to capture a document image. One design goal for the myReader system was an image capture time under 5 seconds. We considered that the ability to quickly scan and analyse a page would be very appealing to users and crucial to the success of the myReader product.

2 Document analysis

After a document image is captured, the greyscale document image is binarised. The skew angle is estimated using a method by Spitz [3], after which the original image is rotated and the rotated image is binarised.

Connected component analysis is applied to locate text and graphics in the binary image. The area Voronoi diagram is computed for all detected components. The Voronoi diagram is examined to find block boundaries, as introduced by Kise [2]. Next, the image is searched to locate horizontal and vertical runs of whitespace, and these are used to split the extracted text blocks.

Characters in each block are connected with their nearest horizontal neighbours to form lines. Unusually tall lines are examined for dropped capitals and split if necessary. Finally, the positions of the characters in each line are examined to detect baselines, and all the characters in the document image are stitched together into a single list for display.

3 Display

myReader offers various display modes, to suit different visual impairments: “live mode”, where a realtime view of the field below the camera is displayed; “image view mode”, where the entire document image is displayed; and several reading modes, shown in Fig. 2: “column mode”, “line mode” and “word mode”.

In all modes, the foreground and background colours can be changed to optimize readability for the user. Image view mode also allows for full colour display. Automatic scrolling is available in image view mode and all reading modes.

Fig. 2. Column mode, with low magnification (left); line mode, with very high magnification (center); word mode, rendering black text on a white background (right).

Column mode reformats the text to display in one long column. As the magnification is increased, fewer characters are displayed on each line. This mode suits users who wish to change the colour or contrast of the reading material but do not require very high magnification levels.

As a large amount of text can be displayed on the screen in column mode, a relatively slow automatic scrolling rate is typically used. This suits the myReader unit's LCD screen as motion blur is not visible at low scroll rates.

Line mode reformats the text to display in a single long horizontal line. For users with visual impairments requiring significant magnification, this format is more comfortable to read than column mode.

In conjunction with automatic scrolling, line mode is especially useful for users with severe loss of peripheral vision or macular degeneration, as the image of the moving text can be understood as it passes over the functional parts of the user's retina.

In word mode, one word is displayed at a time, in the center of the screen. Some users find this mode useful for speed-reading.

4 Results

myReader is able to capture a page, perform document analysis and begin to display reformatted text in 5-10 seconds, depending on the complexity of the page. It is most proficient at handling the complex Manhattan page layouts for which it has been optimized, while glossy magazines, which typically have complex layouts, are the hardest for the system to deal with.

Figure 3 shows some pages that the system deals with well, such as simple text in one or two columns or rectangular textual areas surrounded by pictures. Figure 4 shows examples of pages that are not analysed accurately by myReader, because text overlaps pictures, text and images are not easily separated with rectangular boxes, or there is a wide range of font sizes present on the page. In general, pages that are not handled well do not conform to a complex Manhattan layout. myReader users can still view these pages using image view mode.

Fig. 3. Examples of documents that are easily handled by myReader.

Fig. 4. Examples of documents that are not easily handled by myReader.

During the development of myReader, it was confirmed that the quality of the captured image is very important to the accuracy of document analysis. For instance, an image with poor resolution, or uneven focus or lighting, will be more difficult to analyse correctly. The quality of captured images is generally poorer for a face-up overhead camera-based system than a flatbed scanner, however the overhead camera allows faster page capture and makes live video display possible.

An advantage of performing document layout analysis rather than full OCR is the relative language-independence of the system. Although this has yet to be evaluated fully, the system does a reasonable job of analysing scripts that read from left to right and have separated characters, such as Korean and Japanese.

While there were substantial technical achievements in creating the myReader product, the most important achievement is that it allows many people with low vision to be able to read faster, more easily and for longer than with a conventional video magnifier. A user study by Harrison [1], comparing myReader with a conventional video magnifier, found an 18% increase in reading speed and a substantial increase in physical comfort. Image capture, document analysis and text reformatting are important aids in achieving these advantages.

5 Conclusion

myReader is a practical, portable, camera-based, complete end-to-end document analysis and presentation system that has shown clear benefits for low vision users. The document analysis engine performs well with complex Manhattan-based layouts. Although it has been designed for a low vision reading aid, it could well have other exciting applications.

6 Acknowledgements

Thanks to the rest of the myReader team, and everyone at Humanware who has contributed to the design and development of myReader. Thanks also to Larry Spitz for his ideas and evaluations.

Dedicated to Russell Smith, who tragically passed away in August 2005, but proudly saw myReader through to its release.

7 References

1. C. Harrison. Low-Vision Reading Aids: Reading as a Pleasurable Experience. *Journal of Personal and Ubiquitous Computing* 8(3-4), 2004, pp. 213–220.

2. K. Kise, M. Iwata, and K. Matsumoto. On the Application of Voronoi Diagrams to Page Segmentation. In *Proc. Document Layout Interpretation and its Applications*, Bangalore, 1999.

3. A. Spitz. Skew Determination in CCITT Group 4 Compressed Images, *Symposium on Document Analysis and Information Retrieval*, Las Vegas, Nevada, 1992.